Hausmann, M. (2014). Arts versus science – Academic background implicitly activates gender stereotypes on cognitive abilities with threat raising men's (but lowering women's) performance. *Intelligence* 46, 235-245.

Arts versus Science

– Academic Background Implicitly Activates Gender Stereotypes on Cognitive Abilities With Threat Raising Men's (But Lowering Women's) Performance

Markus Hausmann

Department of Psychology, Durham University, United Kingdom

Address for correspondence:
Markus Hausmann
Department of Psychology, Durham University,
South Road, Durham DH1 3LE, United Kingdom,
Phone: +44-(0)191-33-43286
Fax: +44-(0)191-33-14377
E-mail: Markus.Hausmann@durham.ac.uk

Key words: cognitive sex differences, gender stereotypes, mental rotation, STEM, stereotype threat, verbal fluency

Abstract

There is ongoing debate as to whether "innate" cognitive sex differences contribute to the underrepresentation of women in science and engineering careers. Decades of gender research have revealed good evidence that both biological (e.g. sex hormones) and socio-cultural factors (e.g. gender stereotypes) contribute significantly to cognitive sex differences. Research on gender stereotypes has revealed that priming gender can have adverse or beneficial effects on cognitive performance, depending on whether primed participants appraise the testing situation as threatening or challenging. Several contextual factors have been investigated in this respect. Despite the debate on women in STEM disciplines, however, surprisingly little attention has been paid to academic discipline as a potentially relevant contextual factor. The present study investigated whether gender stereotypes affect cognitive sex differences differently in STEM (chemistry, engineering) and arts (English, philosophy) students. In Experiment 1, male and female arts and science students were tested on two sex-sensitive cognitive tests (mental rotation and verbal fluency) after gender stereotypes were activated. In Experiment 2, arts versus science stereotypes were activated. It was hypothesized that beliefs linked to gender and academic discipline are strongly associated (science = male, arts = female) with similar cognitive effects. Regardless of which identity is primed, it was hypothesized that female arts students would be particularly vulnerable to stereotype threat and would show the lowest performance of all groups in a male cognitive domain (i.e., mental rotation). Due to men's higher confidence in their cognitive abilities, it was hypothesized that primed men would show a performance increase in both spatial (stereotype lift) and verbal abilities (stereotype reactance). The results supported these hypotheses. The two experiments suggest that prompting participants' academic discipline implicitly activated gender stereotypes with considerable negative consequences for women's cognitive test performance. The results also suggest that the well-known sex difference in mental rotation (with men

outperforming women) primarily occurs when negative stereotypes about women's spatial abilities are implicitly primed.

Introduction

The number of women in science, technology, engineering, and mathematics (commonly abbreviated STEM) has significantly increased over the past years. However, women still remain the minority in STEM disciplines and the disparity widens along the educational-vocational continuum, starting at school and gradually increasing during professional (academic) life (Halpern, Benbow, Geary, Gur, Hyde & Gernsbacher, 2007). There is an ongoing debate as to the source of this disparity and there seems to be a tenacious belief that "innate" sex differences in cognitive abilities may partly account for it, or at least that these differences can partly explain why significantly fewer females than males appear at the upper end of higher cognitive abilities that are required in STEM areas (e.g. Summers, 2005, January 14).

Although cognitive performance of both sexes overlap to a large extent, several meta-analyses have revealed that, on average, men perform better than women in specific spatial tasks abilities (Linn & Peterson, 1985; Masters & Sanders, 1993; Voyer, Voyer & Bryden, 1995), particularly in mental rotation (Peters et al., 1995; Vandenberg & Kuse, 1978). In contrast, women perform better, on average, in specific aspects of verbal abilities, such as verbal fluency or verbal memory (Hyde & Linn, 1988; McGlone, 1980). The origins of these cognitive sex differences are still not fully understood (Halpern, 2000). Although there is no doubt that sex hormones and sexual brain dimorphisms contribute to sex differences in specific cognitive abilities, it is also clear that social priming, and especially gender stereotypes can significantly affect men's and women's cognitive performance (e.g. Hausmann, Rosenthal, Schoofs & Jordan, 2009; Hirnstein, Freund, & Hausmann, 2012). In fact, gender stereotypes might be a central variable in environmental influences on sex differences in intelligence (Halpern & LaMay, 2000).

There is broad evidence that social priming can automatically affect individual's behavior, regardless of whether participants are aware of the potential influence of the priming event on their behavior or not (e.g., Bargh, Chen, & Burrows, 1996). An important field of research has investigated the effects of primed stereotypes, usually related to race and gender, on cognitive behavior, such as quantitative skills, spatial cognition, and verbal abilities.

Whether stereotype priming has adverse or beneficial effects on cognitive performance depends on whether participants appraise the testing situation as threatening or challenging. For example, when women were told that a math test consistently shows pronounced sex differences, women's test performance declined, while the same test did not reveal sex differences when introduced as gender-neutral (Spencer, Steele, & Quinn, 1999). Similarly, several studies found that women scored lower in mental rotation tests when they were told that men perform generally better in spatial abilities than women (Moè, 2012; Moè & Pazzaglia, 2006; Wraga, Helt, Jacobs, & Sullivan, 2007). In contrast, cognitive performance improved when participants were confronted with either positive stereotypes about their own group identity or negative stereotypes about the group to which they were compared (e.g., Shih, Pittinsky, & Ambady, 1999; Walton & Cohen, 2003). For example, women showed enhanced performance in mental rotation tests, when they were told these tests measure perspective-taking abilities in which they were superior to men (Moè, 2009; Heil, Jansen, Quaiser-Pohl, & Neuburger, 2012; Wraga et al., 2007; Wraga, Duncan, Jacobs, Helt, & Church, 2006).

Adverse and beneficial effects of stereotype priming even occurred when the priming cue to group identity was less salient. Shih, Pittanski and Ambady (1999) investigated the effects of implicit stereotype priming in Asian American female undergraduates. The researchers primed negative stereotypes related to gender (e.g., women have inferior quantitative skills compared with men) and positive stereotypes related to ethnic identity (e.g., Asians have superior quantitative skills compared to other cultures) by asking participants whether they preferred single-sex floors in their college, and whether their grandparents spoke any language other than English. The results revealed that participants' quantitative skills were altered in the direction predicted by the stereotype associated with gender and ethnic identity.

However, participants' cognitive performance after stereotype priming does not always follow the direction predicted by the stereotype. Many studies have struggled to replicate the 'classic' stereotype threat or reported effects opposite to the direction predicted by the stereotype (e.g., Picho, Rodriquez, & Finnie, 2013, for a review). The diverse effects of stereotype priming can be summarized as follows: If individuals are afraid of confirming a negative stereotype, cognitive performance can decline – a phenomenon called 'stereotype threat' (Steele & Aronson, 1995). In contrast, if confronted with a positive stereotype about the individual's group identity, cognitive performance can improve slightly ('stereotype lift'; Walton & Cohen, 2003) or significantly ('stereotype boost'; Shih, Pittinsky, & Ambady, 1999; Shih, Ambady, Richeson, Fujita, & Gray, 2002). Cognitive performance can also improve when confronted with a negative stereotype about an out-group ('stereotype susceptibility', Walton & Cohen, 2003), or when a negative stereotype about the in-group is appraised as challenging–termed 'stereotype reactance' (Kray, Thomason, & Galinsky, 2001). Stereotype reactance was found especially when stereotypes were explicitly primed. "When a negative stereotype is blatantly and explicitly activated, it might be perceived by the test taker as a limit to their freedom and ability to perform, thereby ironically invoking behavior that is inconsistent with the stereotype" (Nguyen & Ryan, 2008, p. 1315).

Underlining the highly situational character of stereotype threat (Steele, 1997), it has been shown that degree and direction of stereotype priming effects can be mediated by various psychological factors (e.g., gender identification, Schmader, 2002; stigma consciousness, Brown & Pinel, 2003) and contextual factors (e.g., sex composition, Inzlicht & Ben-Zeev, 2003; Hirnstein, Andrews, & Hausmann, 2014; Murphy, Steele, & Gross, 2007; type of priming, Bargh, 1997; test difficulty, Inzlicht & Ben-Zeev, 2003; domain identification, Cadinu et al., 2003; academic domain (STEM vs. Non-STEM, Crisp, Bache, & Maitner, 2009; Werhun, 2007). Two recent meta-analyses estimated the effect sizes of these contextual factors on stereotype threat and found that on average, women under stereotype threat performed nearly a quarter of a standard deviation (d = -0.24, Picho, Rodriquez, & Finnie, 2013; d = -0.21, Nguyen & Ryan, 2008) below their non-stereotyped counterparts on math tests. While this effect size was only marginally affected by, for example, sex composition (i.e., same-sex groups: d = -0.22, mixed-sex groups: d = -0.26) and type of priming (implicit: d = -0.28; explicit: d = -0.23), Picho et al. (2013) revealed that academic domain had differential stereotype threat effects on quantitative skills in STEM (d = 0.06) vs. non-STEM undergraduate students (d = -0.25). However, only a very few studies have looked at academic domain as a mediating factor. This is surprising given the ongoing debate on the underrepresentation of women in science disciplines, and attempts to enhance women's performance in STEM disciplines. One of these studies (Crisp et al., 2009) investigated the effect of negative stereotype priming on quantitative skills in 39 female psychology and engineering majors. This study found that whereas female undergraduate psychology students exhibited a 'classic' stereotype threat, female engineering majors showed a significantly improved performance. The authors ruled out a more generalized stereotype reactance because both groups received identical explicit threat instructions (i.e., "on this test we shall be comparing the performance of males and females'), but only women engineers reacted positively. In addition, the authors speculated that stereotyped individuals who are good at math (i.e., female engineers) are less susceptible to stereotype threat than individuals who are less good at math (i.e., psychology students) because of their coping resources developed as a result of experience and previous success. However, the authors also discussed the possibility that "individuals who have experience in counter-stereotypic domains have an alternative social identity that they are able to shift to in the face of threatening intergroup comparisons […] because they have access to an alternative dimension of self-categorization that affords them to a domain-relevant positive performance stereotype" (Crisp et al., 2009, p. 180).

The present study aimed to investigate the contextual influence of academic domain on gender stereotyping in tasks of mental rotation and verbal fluency. These respectively 'male' and 'female' cognitive domains were used for three reasons. First, although cognitive abilities of the sexes overlap to a large degree (McKeever, 1995), meta-analyses have shown mental rotation and verbal fluency to be particularly sensitive to sex: men outperform women in mental rotation by about 0.6 SD units (Linn & Peterson, 1985; Voyer, Voyer, & Bryden, 1995) and women outperform men in verbal fluency by about 0.3 SD units (Hyde & Linn, 1988). Second, these cognitive abilities are differentially involved in science and arts disciplines. Mental rotation has been considered to be a fundamental skill for success in mathematics and science occupations and degrees (e.g., Linn & Petersen, 1986; Lubinski, 2010; Nazareth, Herrera, & Pruden, 2013; Sherman, 1967; Wai, Lubinski, & Benbow, 2009). Science students deal with spatially relevant material such as geometry (Baenninger & Newcombe, 1989) and spatial construction tasks more frequently than students enrolled in humanities and social sciences (Jordan, 2010). Participants enrolled in physical science majors (e.g., chemistry) outperform arts students in mental rotation (Peters, Laeng, Latham, Jackson, Zaiyouna, & Richardson, 1995). Similarly, Casey and Brabeck (1989) found that spatial experience (in combination with other biological factors) in women who had majored in math or science eliminated sex differences in mental rotation (see also Richardson, 1994). Verbal abilities, in which women usually excel, are assumed to be relevant in all academic areas (Halpern et al., 2007). However, the finding that arts students outperformed science students in linguistic skills and essay writing (North, 2005) suggests that, because more women than men take part in arts subjects, academic discipline and experience may contribute to sex differences in verbal abilities. Finally, both mental rotation and verbal fluency are affected by gender stereotypes. Women scored lower in mental rotation when confronted with stereotypes about men's superior spatial abilities (Wraga, Helt, Jacobs, & Sullivan, 2007; Moè & Pazzaglia, 2006), whereas they scored higher in mental rotation when the test was introduced as a measure of perspective-taking abilities in which women are superior to men (Moè, 2009; Wraga, Duncan, Jacobs, Helt, & Church, 2006; Wraga, Helt, Jacobs, & Sillivan, 2007). A recent study showed that gender stereotyping can also affect verbal fluency in both men and women (Hirnstein et al., 2012). However, this study found that verbal fluency in men and women increased when gender stereotypes were activated, indicating a stereotype lift and/or susceptibility in women, and a stereotype reactance in men.

For mental rotation (Experiment 1), it was predicted that science students would outperform arts students because of their experience with 3D visualization (Hypothesis 1). Men were predicted to outperform women (Hypothesis 2), and this cognitive sex difference was hypothesized to increase when gender-stereotypes were primed, particularly in individuals in academic disciplines that correspond to the gender stereotype. In other words, it was hypothesized that female arts students would be especially prone to stereotype threat when performing mental rotation (Hypothesis 3). Moreover, and in line with Crisp et al. (2009), it was hypothesized that gender stereotype-primed female science students perform better on mental rotation than their non-stereotyped counterparts (Hypothesis 4), because of their (a) experience with 3D visualization, and/or (b) ability to access an alternative dimension of self-categorization that affords them a domain-relevant positive performance stereotype (Crisp et al., 2009). For verbal fluency (Experiment 1), it was predicted that arts students would outperform science students (Hypothesis 5), and that women would outperform men, at least when no gender stereotypes were primed (Hypothesis 6). In contrast to primed women performing the mental rotation test, and in line with previous studies (Hirnstein, et al., 2012), it was also predicted that men would be less susceptible to stereotype threat, and would show reactant effects when performing the verbal fluency task (Hypothesis 7).

The heated debate on STEM and "innate" cognitive sex differences implicitly links men's and women's cognitive abilities to cognitive profiles related to science and arts disciplines, respectively. This would predict that prompting participants' academic discipline (Experiment 2) would implicitly activate gender stereotypes, leading to similar results as predicted in Experiment 1.

Experiment 1 – Gender Stereotyping

Methods

Participants. Eighty-eight undergraduate students from four academic departments at Durham University (44 men, 44 women) participated in this experiment (mean age: women: 20.21 years, men: 20.79 years). Forty-four participants were studying for arts degrees (22 English, 22 philosophy) and 44 for science degrees (22 chemistry, 22 engineering), with equal numbers of men and women in each group. The minimum entry requirements for all Durham students in these four academic disciplines is three A grades or better. Academic domains (i.e., arts and science) were not further divided into majors in the statistical analyses, because of relatively small sample sizes. Participants were randomly assigned to either the experimental (gender-stereotyped) or control group. All participants were volunteers recruited by announcements within the university and they were tested individually.

Procedure and materials

Gender-stereotype questions. To measure participants gender stereotypes on mental rotation and verbal fluency, and thereby to implicitly activate their gender stereotypes, stereotypes were measured before cognitive testing (for details Hausmann et al., 2009; adopted from Halpern & Tan, 2001). Participants in the experimental group were told to imagine that they were about to meet a person who they had never met before and they were required to estimate the probability that the individual was 'male' or 'female' given that this person "… *can imagine abstract objects and rotate them mentally in all direction*" (mental rotation stereotype) and "… *can generate many words beginning with the same letter in one minute*" (word fluency stereotype). Two columns were aligned next to each item (labelled *male* and *female*) and participants entered a number that corresponded to their probability estimate, with the two estimates summing to 100. A probability estimate of 50% for 'male' and 50% for 'female' would indicate that the participant perceive no stereotypical gender differences in this particular cognitive domain. Participants in the control group estimated the probability that the same cognitive abilities would be more or less associated with being "North American" or "European".

Self-ratings. Self-rating were also measured (Hausmann et al., 2009, for details), using a seven-point scale, with 1 = *not at all descriptive of me* to 7 = *highly descriptive of me*. Self-ratings were measured to assess the possibility that participants may believe that an ability (i.e., performance in mental and word fluency) is generally associated with one sex or the other, but that as an individual, he or she is an exception to these stereotypes.

Mental Rotation Test. Spatial ability was assessed with the Revised Vandenberg & Kuse Mental Rotations Tests– Version MRT-A (Peters et al., 1995) which involves 2D drawings of 3D cube figures (Shepard & Metzler, 1971). The mental rotation test contains two sets of 12 items. For each set, participants have a time limit of three minutes, with a three-minute break between the two sets. Each item consists of a target figure on the left and four stimulus figures on the right. Two of these stimulus figures are rotated versions of the target figure, and two of the stimulus figures cannot be matched to the target figure. One point is given if both matching stimulus figures are correctly identified. The maximum score in this test is 24 points.

Word Fluency Test. Verbal ability was assessed by the word fluency test, a subtest of the Leistungsprüfsystem (LPS; Horn, 1962). Participants successively receive two letters (L and P) and have one minute per letter to generate as many words (excluding names) as possible in a written format. One point is given for each correct word.

Results

Gender stereotypes. To investigate the strength of mental-rotation stereotypes related to gender in the current sample, mean probability estimates of being male for the experimental group were entered to a 2 × 2 ANOVA with sex and academic discipline (arts degrees, science degrees) as between-subject factors. The analysis revealed a significant main effect of academic discipline, $F(1, 40) = 8.05$, $p = .007$, $\eta_p^2 = .17$, indicating that science students (M = 67.73%, SD = 11.93%) were more convinced than arts students (M = 57.27%, SD = 12.88%) that someone good in mental rotation was likely to be male. No other main effect or interaction approached significance, all $F \leq 2.57$, ns. For stereotypes related to word fluency, the analysis also revealed a significant main effect of academic discipline, $F(1, 40) = 4.10$, $p < .05$, $\eta_p^2 = .09$, showing that arts students (M = 41.14%, SD = 9.50%) were less convinced than science students (M = 46.14%, SD = 6.16%) that someone good in verbal fluency was likely to be male. Again, no other effect approached significance, all $F \leq 0.14$, ns (see Table 1).

Self-ratings. A 2 × 2 × 2 ANOVA of self-ratings on mental rotation with academic discipline (arts degrees, science degrees), sex, and condition (stereotype priming, control) as between-subject factors revealed a significant main effect of sex, $F(1, 80) = 6.55$, $p = .012$, $\eta_p^2 = .08$, showing that men ($M = 4.59$, SD = 1.45) were more self-confident in their mental rotation skills than women ($M = 3.68$, SD = 1.91). Also, self-ratings by science students ($M = 4.45$, SD = 1.72) were slightly higher than those by arts students ($M = 3.82$, SD = 1.74), $F(1, 80) = 3.21$, $p = .077$, $\eta_p^2 = .04$). No other effect was significant, all $F \leq 2.36$, ns. For verbal fluency, the main effect of discipline was significant, $F(1, 80) = 18.22$, $p < .001$, $\eta_p^2 = .19$, with greater self-confidence in arts students ($M = 5.23$, SD = 1.48) than science students ($M = 3.98$, SD = 1.28). No other effect was significant, all $F \leq 3.77$, ns (see Table 1).

Table 1. Gender stereotypes (i.e., mean probability estimates of being male) and self-ratings for mental rotation and verbal fluency in men and women of the experimental group before cognitive testing, and after stereotyping gender (Experiment 1). Standard deviations are shown in brackets.

Ability	**Probability estimates (male/science)**				**Self-ratings (0-7)**			
	Science degrees		**Arts degrees**		**Science degrees**		**Arts degrees**	
	Men	**Women**	**Male**	**Women**	**Male**	**Women**	**Male**	**Women**
Mental rotation	**69.1 (14.3)**	**66.4 (9.51)**	52.7 (5.18)	**61.8 (16.6)**	**4.91 (1.22)**	4.27 (2.24)	**5.18 (0.40)**	**3.00 (1.41)**
Word fluency	46.8 (7.17)	**45.5 (5.22)**	**41.4 (8.97)**	**40.9 (10.4)**	4.09 (1.30)	4.27 (0.65)	**5.63 (0.92)**	**5.55 (1.29)**

Note: Bold probability estimates and self-ratings indicate significant differences from 50% and test score 4, respectively (p < .05). Probability estimates above 50% indicate stereotypes favoring men/science. Probability estimates below 50% indicate stereotypes favoring women/arts.

Cognitive performance. To investigate the effect of academic discipline, sex, and stereotype priming on mental rotation performance, a 2 × 2 × 2 ANOVA was calculated. The analysis revealed the expected main effects for academic discipline and sex. As predicted (Hypothesis 1), science students (M = 14.00, SD = 4.40) outperformed arts students (M = 9.27, SD = 4.97), $F(1, 80) = 30.67$, $p < .001$, $\eta_p^2 = .28$, and men (M = 13.86, SD = 4.59) outperformed women (M = 9.41, SD = 4.92), $F(1, 80) = 27.23$, $p < .001$, $\eta_p^2 = .25$ (Hypothesis 2), both with large effect sizes. Moreover, the 3-way interaction was significant, $F(1, 80) = 6.00$, $p = .016$, $\eta_p^2 = .07$. To analyze the nature of the 3-way interaction, two subsequent 2 × 2 ANOVAs were performed, one for each priming condition. As expected (Hypothesis 3), the ANOVA revealed a significant discipline × sex interaction only in the gender-stereotype condition, $F(1, 40) = 8.27$, $p = .006$, $\eta_p^2 = .17$, not in the gender-neutral control condition, $F(1, 80) = 0.77$, ns, $\eta_p^2 = .02$ (see Figure 1).

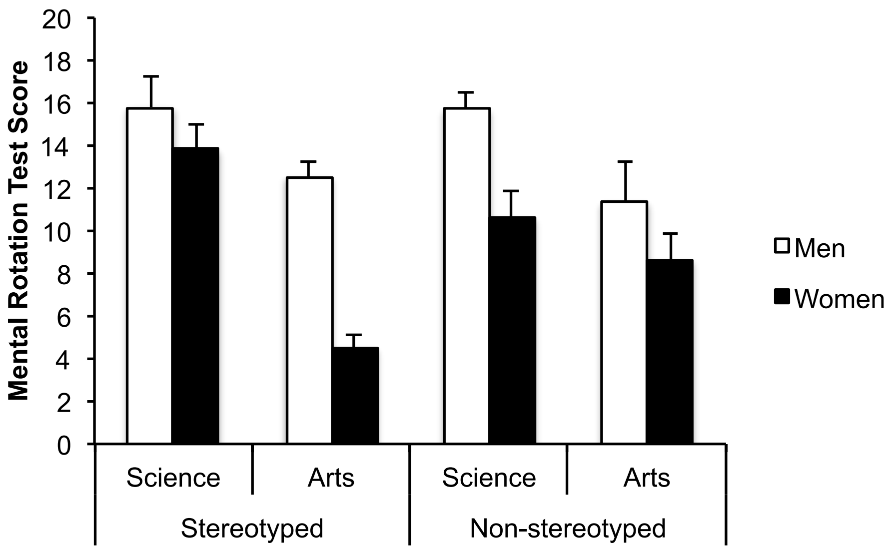

Figure 1. Significant second-order interaction (Experiment 1) between academic discipline (science degrees, arts degrees) and sex in mental rotation performance (means ± standard error means) after priming gender (stereotyped, left panel) and no prime (non-stereotyped, right panel).

These findings indicate that the sex difference in mental rotation after priming largely depended on academic discipline, accounting for 17% of explained variance, as compared to less than 2% in the non-primed condition. To investigate whether female arts students were particularly susceptible to stereotype threat (Hypothesis 3), independent t-tests on gender-stereotype conditions were performed. As expected, primed female arts students not only performed significantly worse than primed male arts students, $t(20) = 8.97$, $p < .001$, d = 3.83, but also worse than their non-primed counterparts, $t(20) = 3.06$, $p = .006$, d = 1.10. Finally, an independent t-test revealed that, as expected (Hypothesis 4), stereotype-primed female science students performed better on mental rotation than their non-stereotyped counterparts, $t(20) = 1.79$, $p = .044$, one-tailed, d = 0.73. Overall, the results indicated large effect sizes for gender priming. No other effect approached significance, all $F \leq 3.28$ (see Table 2).

For word fluency, the same 2 × 2 × 2 ANOVA revealed the expected main effect of discipline, $F(1, 80) = 46.40$, $p < .001$, $\eta_p^2 = .37$, with arts students ($M = 20.73$, SD = 5.52) outperforming science students ($M = 13.91$, SD = 4.60) (Hypothesis 5). In addition, the main effect of condition was significant, $F(1, 80) = 10.99$, $p = .001$, $\eta_p^2 = .12$. Stereotyped participants ($M = 18.98$, SD = 5.96) generally outperformed non-primed controls ($M = 15.66$, SD = 5.85). The expected main effect of sex on word fluency was not significant (Hypothesis 6), $F(1, 80) = 0.74$, ns, and depended on condition, as indicated by the significant sex × condition interaction, $F(1, 80) = 7.42$, $p = .008$, $\eta_p^2 = .12$. Independent t-tests revealed that women, independent of their academic discipline, outperformed men in the non-primed control condition, $t(42) = 2.12$, $p = .02$, d = .90, but not in the gender-stereotype condition, $t(42) = 1.04$, $p = .31$, d = 0.26, which generally enhanced word fluency performance in both sexes (Hypothesis 7). No other effect approached significance, all $F \leq 1.19$, ns (see Table 2).

Table 2. Cognitive performance (mean ± standard deviation) in mental rotation (MRT) and verbal fluency (WF) after stereotyping gender (Experiment 1) according to condition (stereotyped, non-stereotyped), sex, and academic discipline (science degrees, arts degrees).

		Stereotyped group		Non-stereotyped Controls	
Task	**Discipline**	**Men**	**Women**	**Men**	**Women**
MRT	Science	15.82 ± 4.73	13.82 ± 4.09	15.73 ± 2.61	10.63 ± 4.23
	Arts	12.55 ± 2.25	4.55 ± 1.92	11.36 ± 6.30	8.64 ± 4.01
WF	Science	16.91 ± 3.30	13.73 ± 4.65	10.27 ± 3.98	14.73 ± 4.17
	Arts	22.91 ± 3.56	22.36 ± 6.53	17.45 ± 5.56	20.18 ± 4.94

Predicting cognitive performance by gender stereotypes and self-ratings.

To investigate whether the degree of gender-stereotypes and/or self-confidence in a particular cognitive test predicted cognitive performance, two multiple linear regressions were performed on all participants in the gender-stereotype condition. Gender-stereotype probability estimate and self-rating score were entered as predictors and performance on the two cognitive tests were outcome variables. For the mental rotation test, multiple regression revealed that only self-ratings significantly predicted mental rotation performance, $\beta = .43$, $p = .004$; the higher participants' self-confidence, the better the performance. The strength of gender stereotypes (i.e., a stronger belief that men are more likely to perform better in mental rotation than women) did not predict mental rotation performance. For the word fluency test, the multiple regression revealed that both self-ratings, $\beta = .56$, $p < .001$, and gender-stereotype probability estimates, $\beta = -.38$, $p = .003$, predicted word fluency accounting for 42.4% of variance. The higher the word-fluency self-rating, and the more pronounced the word-fluency gender stereotype ('being female'), the higher the word-fluency score in men and women (see Table 3).

Table 3. Multiple linear regressions (standardized beta coefficients) for probability estimates and self-ratings as predictors of cognitive performance in the mental rotation test (MRT) and word fluency (WF) after stereotyping gender (Experiment 1) and academic discipline (Experiment 2). Determination coefficients (R^2) and significances (*p*) indicate the goodness-of-fit for the regression model.

Task	Experiment	Probability estimates	Self-ratings	R^2	*p*
MRT	1	.09	**.43***	.21	.009
	2	.00	**.59***	.31	< .001
WF	1	**-.38***	**.56***	.42	< .001
	2	-.14	**.44***	.12	.023

Note: Bold probability estimates and self-ratings contribute significantly (*p < .01) to the regression model.

Summary. Taken together, all participants held pronounced gender stereotypes for mental rotation and word fluency, although degree and direction of participants' gender stereotypes varied with academic discipline rather than sex. Both academic groups showed pronounced cognitive differences in the expected direction (Hypotheses 1 and 5). Robust sex differences were only found for mental rotation (i.e., men outperformed women; Hypothesis 2). Women outperformed men in word fluency only in non-primed groups (Hypothesis 6). Most importantly, a robust sex difference in mental rotation was mainly driven by a stereotype threat effect in female arts students (Hypothesis 3), suggesting that this group was particularly vulnerable to stereotype threat in a cognitive domain that is perceived as being associated with science and being male. Mental rotation performance in gender-stereotyped female science students was in the male performance range, indicative of stereotype reactance (Hypothesis 4). For word fluency, gender stereotyping resulted in a stereotype boost, particularly in men, that was independent of academic discipline (Hypothesis 7), thereby eliminating the sex difference in word fluency. For both tasks, cognitive performance was more related to self-ratings than gender stereotypes, suggesting that self-confidence is more predictive of cognitive effects after gender stereotyping than the degree of stereotyping.

Experiment 2: Academic Discipline Stereotyping

Method

Participants. Ninety-six undergraduate students (48 men, 48 women) from Durham University, UK, participated in Experiment 2 (mean age: women: 20.10 years, men: 21.31 years). Forty-eight participants were studying for arts degrees (24 philosophy, 24 English) and 48 for science degrees (24 chemistry, 24 engineering) with equal numbers of men and women. As in Experiment 1, academic domains (i.e., arts and science) were not further divided into majors due to relatively small sample sizes. Participants were again randomly assigned to either the experimental (academic-discipline stereotyped) or control group. All participants were recruited by announcements on a voluntary basis and they were tested individually.

Procedure and materials

The stereotyping, self-ratings and cognitive tests were identical to Experiment 1 with one alteration. Participants in the experimental group were told to imagine that they were about to meet a person who they had never met before and they were required to estimate the probability that the individual was a *'science'* or *'arts'* student. The analysis focused again on the two items that directly referred to mental rotation and verbal fluency.

Results and Discussion

Discipline-stereotype questions. To investigate the strength of the mental-rotation stereotype related to academic discipline, a 2 × 2 ANOVA of the mean probability estimates of being a science rather than arts student for the stereotype-primed groups was performed with academic discipline (science degrees, arts degrees) and sex as between-subject factors. The ANOVA revealed no significant effects, all $F \leq 1.17$, ns, suggesting that mental-rotation stereotypes related to discipline were consistent across all groups (i.e., science students perform better on mental rotation than arts students). The same ANOVA for the word-fluency stereotype found only the academic discipline × sex interaction to be significant, $F(1, 44) = 6.69$, $p = .013$, $\eta_p^2 = .13$. Posthoc t-tests revealed that female science students had stronger beliefs than female arts students that someone doing well in word fluency must be an arts student, $t(22) = 2.11$, $p < .05$. No other effect was significant, all $F \leq 0.27$, ns. These findings were similar to Experiment 1 (see Table 4).

Table 4. Academic stereotypes (i.e., mean probability estimates of being a science student) and self-ratings for mental rotation and verbal fluency in men and women of the experimental group before cognitive testing, and after stereotyping academic background (Experiment 2). Standard deviations are shown in brackets.

Ability	Probability estimates (science)				Self-ratings (0-7)			
	Science degrees		Arts degrees		Science degrees		Arts degrees	
	Men	Women	Male	Women	Male	Women	Male	Women
Mental rotation	**75.4 (11.2)**	**65.4 (19.2)**	**68.8 (20.6)**	**68.8 (10.5)**	**5.25 (1.36)**	4.75 (1.42)	3.50 (1.38)	**2.75 (1.60)**
Word fluency	**42.9 (8.65)**	**32.1 (16.2)**	**35.4 (14.7)**	**43.3 (8.88)**	3.83 (1.53)	**2.58 (1.51)**	4.08 (1.62)	4.17 (1.34)

Note: Bold probability estimates and self-ratings indicate significant differences from 50% and test score 4, respectively (p < .05). Probability estimates above 50% indicate stereotypes favoring men/science. Probability estimates below 50% indicate stereotypes favoring women/arts.

Self-ratings. For mental rotation, as expected, self ratings by science students ($M = 4.92$, SD = 1.40) were higher than arts students ($M = 3.19$, SD = 1.39), $F(1, 88) = 37.39$, $p < .001$, $\eta_p^2 = .30$. Similarly, men ($M = 4.42$, SD = 1.41) gave higher self ratings than women ($M = 3.68$, SD = 1.78), $F(1, 88) = 6.39$, $p = .012$, $\eta_p^2 = .07$. No other effect approached significance, all $F \leq 0.26$, ns. For word fluency, only the main effect of academic discipline was significant, $F(1, 88) = 7.58$, $p = .007$, $\eta_p^2 = .08$, with higher self rating by arts ($M = 4.15$, SD = 1.61) than science students (M = 3.23, SD = 1.61). No other effect was significant, all $F \leq 1.57$, ns (see Table 4).

Cognitive performance. To investigate the influence of academic discipline, sex, and discipline priming on mental-rotation performance, a 2 × 2 × 2 ANOVA was calculated. The analysis replicated the expected main effects for academic discipline and sex observed in Experiment 1. Again, science students (M = 13.79, SD = 4.92) outperformed arts students (M = 7.96, SD = 3.92), $F(1, 88) = 56.90$, $p < .001$, $\eta_p^2 = .39$ (Hypothesis 1), and men (M = 12.60, SD = 5.77) outperformed women (M = 9.15, SD = 4.19), $F(1, 88) = 20.00$, $p < .001$, $\eta_p^2 = .19$ (Hypothesis 2). In contrast to Experiment 1, the sex × condition interaction was significant, $F(1, 88) = 5.88$, $p < .017$, $\eta_p^2 = .06$. Post hoc tests explored this interaction further and revealed that men outperformed women only in the stereotype condition, $t(46) = 3.36$, p = .002, d = 1.37, suggesting a stereotype lift in discipline-primed men. Also, as expected, women who had experienced discipline priming showed a depressed score on mental rotation compared to control women, $t(46) = 3.03$, p = .004, d = 1.24. Finally, and in line with Experiment 1, the 3-way interaction was significant, $F(1, 88) = 4.88$, $p = .03$, $\eta_p^2 = .05$. Again, two subsequent 2 × 2 ANOVAs were performed, one for each priming condition. As expected, and in line with Experiment 1, the ANOVA revealed a significant academic discipline × sex interaction only in the stereotype condition, $F(1, 44) = 7.91$, $p = .007$, η_p^2 = .15, not in the control condition, $F(1, 44) = 0.09$, ns, (see Figure 2).

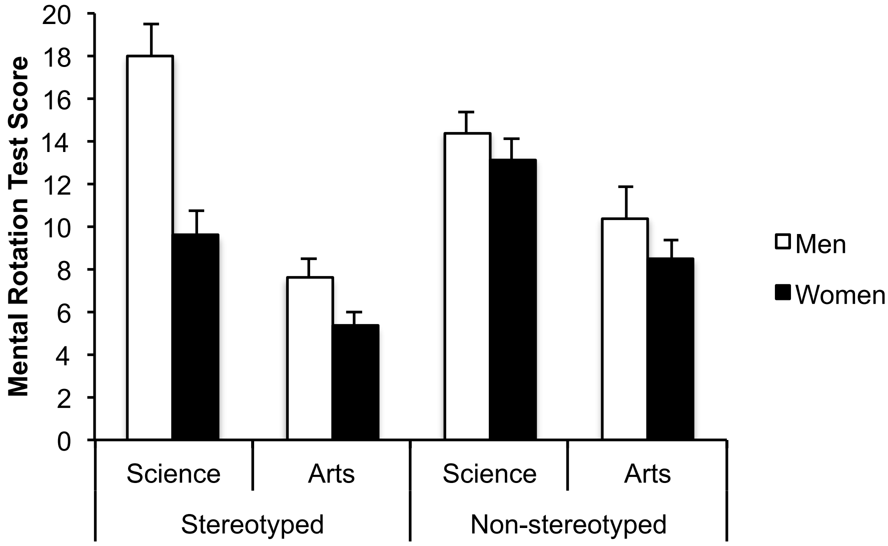

Figure 2. Significant second-order interaction (Experiment 2) between academic discipline (science degrees, arts degrees) and sex in mental rotation performance (means ± standard error means) after priming academic discipline (stereotyped, left panel) and no prime (non-stereotyped, right panel).

This finding supports the previous conclusion that the sex difference in mental-rotation performance after priming depends on academic discipline (Hypothesis 3), accounting for 15% of variance (similar to 17% found in Experiment 1). Again, primed female arts students showed significantly lower mental rotation performance than their primed male counterparts (Hypothesis 3), $t(22) = 2.02$, $p = .03$, one-tailed, d = 0.78. However, in contrast to Experiment 1 which revealed the expected stereotype reactance in primed female *science* students (Hypothesis 4), in Experiment 2 this group performed worse than primed male science students, $t(22) = 4.46$, $p < .001$, d = 1.35, and also worse than their non-primed counterparts, $t(22) = 2.43$, $p = .02$, d = 0.90 (see Table 5). In other words, Experiment 2 revealed stereotype threat in women from both academic disciplines, suggesting that academic background priming implicitly primed negative stereotypes related to gender (e.g., women have inferior quantitative skills compared with men). This finding suggests that experience with 3D visualization, and/or the ability to access an alternative dimension of self-categorization did not necessarily lead to a domain-relevant positive performance stereotype (Crisp et al., 2009), and consequently better performance. No other effect was significant, all $F \leq 3.76$, ns.

Table 5. Cognitive performance (mean ± standard deviation) in mental rotation (MRT) and verbal fluency (WF) after stereotyping academic discipline (Experiment 2) according to condition (stereotyped, non-stereotyped), sex, and academic discipline (science degrees, arts degrees).

		Stereotyped group		**Non-stereotyped Controls**	
Task	**Discipline**	**Men**	**Women**	**Men**	**Women**
MRT	Science	18.00 ± 5.17	9.58 ± 4.01	14.42 ± 3.32	13.17 ± 3.16
	Arts	7.58 ± 3.12)	5.33 ± 2.70	10.42 ± 5.16	8.50 ± 3.09
WF	Science	18.58 ± 4.81	16.58 ± 6.22	15.25 ± 5.41	15.17 ± 3.46
	Arts	20.00 ± 7.02	21.33 ± 5.26	9.64-15.27	17.08 ± 3.75

For word fluency, the ANOVA revealed only a significant condition effect, $F(1, 88) = 14.44$, $p < .001$, $\eta_p^2 = .14$) in which stereotyped participants of both sexes ($M = 19.13$, SD = 5.97) outperformed controls ($M = 15.17$, SD = 4.27), consistent with the idea that primed men are generally less susceptible to stereotype threat (Hypothesis 7), and are even reactant to word fluency stereotype. No other effect was significant, all $F \leq 3.10$, ns (see Table 3). This included the main effect of academic discipline and sex (Hypothesis 5 and 6), suggesting that (a) arts and science students performed similarly in word fluency when discipline, not gender, was primed, and (b) word fluency was generally less sex-sensitive than the mental rotation test (in line with Experiment 1).

Predicting cognitive performance by academic stereotypes and self-ratings.

To investigate the influence of discipline stereotypes and self-confidence on test performance, multiple regressions were performed. Similar to Experiment 1, only self-ratings were significantly related to mental-rotation performance, $\beta = .59$, $p < .001$. For word fluency also, the regression revealed that only self-ratings were significantly (and positively) related to performance, $\beta = .44$, $p = .007$ (see Table 3). In sum, similar to Experiment 1, self-confidence was a better predictor of task performance than the extent to which participants believed that men and women differ in their task abilities.

4. General Discussion

In these studies, I examined the effects of stereotyping gender and academic discipline (science degrees versus arts degrees) on cognitive performance in a 'male' domain (mental rotation) and a 'female' domain (verbal fluency) according to academic discipline (science versus arts). Overall, the results revealed the expected cognitive differences between academic groups. Science students outperformed arts students in mental rotation, whereas the opposite was the case for verbal fluency (Hypothesis 1 and Hypothesis 5). In line with previous findings (Linn & Petersen 1985; Voyer et al., 1995), men outperformed women in mental rotation (Hypothesis 2). Moreover, the results indicated that, as expected, gender-stereotypes affected mental rotation performance mainly in female arts students for which (two) negative stereotypes exist (Hypothesis 3). However, gender stereotyping can have opposite effects, at least in women, and depends on whether stereotypic or counter-stereotypic group identity associated with gender or academic discipline was activated. Thus, the enhanced mental rotation test performance in science women (Experiment 1) can be explained by stereotype reactance when a counter-stereotypic group identity associated with academic discipline was implicitly activated (Hypothesis 4). Together with the inferior performance in arts women (i.e., stereotype threat), this resulted in a large difference in mental-rotation performance between these two groups of women (d = 3.66). In contrast, men's cognitive performance generally improved when stereotyped, regardless of the positive or negative nature of the primed stereotype (Hypothesis 7). This may also partly explain why the current study found the expected sex difference in word fluency (Hypothesis 6) only in the non-primed condition of Experiment 1. In line with previous meta-analyses (Hyde & Linn, 1988; Linn & Peterson, 1985; Voyer, Voyer, & Bryden, 1995), the current study revealed verbal fluency to be less sex-sensitive than mental rotation. Overall, the results suggest that (a) the effects of gender stereotyping are sex-specific, and (b) women are generally more susceptible to stereotype threat than men.

It is important to note that both experiments revealed similar results. That is, stereotyping affected cognitive performance regardless of whether gender or academic discipline was primed. This suggests that the beliefs linked to both group identities are strongly associated (science = male, arts = female). Consequently, priming someone's academic discipline can implicitly activate gender stereotypes with similar cognitive effects. The reason why mental rotation was especially susceptible to stereotype threat might be related to the fact that mental rotation is cognitively more demanding than word fluency (see Hausmann et al., 2009), and it has been suggested that more difficult tasks lead to higher arousal which, if appraised as threat, results in a performance decrease, i.e. stereotype threat (Ben-Zeev, Fein, & Inzlicht, 2005; O'Brien & Grandall, 2003). Related to this, tasks involving a high working memory load, which seems to be the case for mental rotation (Hyun & Luck, 2007; Zimmer, 2008), are more prone to stereotype threat (Schmader & Johns, 2003; Schmader, Johns, & Forbes, 2005). A recent study also found that distinct executive functions were critically affected when stereotype threat led women to underperform at math tasks (Rydell, Van Loo, & Boucher, 2013).

However, there were two main differences between the experiments. Although both experiments revealed stereotype threat on mental-rotation performance in female arts students, female science students only showed a stereotype threat in Experiment 2. One possible explanation for this effect is that two group identities with different implications for task performance were primed in female science students, i.e., science student identity in Experiment 1 and gender identity in Experiment 2, resulting in stereotype lift and stereotype threat, respectively. An alternative explanation might be that academic priming activated stereotypes of both academic discipline and gender, and so doubled the pressure to fulfill the high expectations related to the science stereotype in a ‘male’ cognitive domain (compared to when only gender-stereotypes were primed).

Pronounced gender-stereotypes of spatial and verbal abilities were found in all groups, independent of sex and academic group. However, the large effect sizes for the belief that men outperform women in spatial abilities, and that women perform better than men in verbal abilities was greater than actual cognitive sex differences for which effect sizes range from (very) small to medium (Hyde & Linn, 1988; Linn & Peterson, 1985; Voyer et al., 1995). These ‘overemphasized stereotypes’ in specific cognitive abilities have been reported before (Swim, 1994) and may explain why the strength of self-reported stereotypes for gender and discipline was not directly related to cognitive performance in mental rotation and verbal fluency. Gender stereotypes were especially overemphasized in those students for whom a positive academic-discipline stereotype existed, thereby increasing their self-confidence, and eventually enhancing cognitive performance in these groups. In fact, confidence in one’s own cognitive abilities, as measured by self-ratings before cognitive testing, was positively related to mental rotation and word fluency performance in all stereotyped groups. The sex difference in self-ratings explained the ‘stereotype lift’ found primarily in stereotyped male science and arts students. Previous research has also shown that confidence partially mediates sex differences in mental rotation (Estes & Felker, 2012). Stereotyped men may have appraised the test situation as challenging rather than threatening (Hirnstein et al., 2012; Hausmann et al., 2009). The increase in men’s performance (stereotype reactance and/or stereotype lift) occurred regardless of the cognitive domain, and in response to both positive and negative stereotypes. Recent research suggests that this finding, and the observation that men are more likely to interpret the testing situation as challenging, relates to testosterone levels (Hausmann et al., 2009; Josephs, Newman, Brown, & Beer, 2003). In other words, the reduced susceptibility to stereotype threat in men might be partly biological in nature.

The generally lower self confidence reported by women, and its relationship with poorer cognitive performance, suggests higher anxiety levels in women in ‘male’ cognitive domains, especially in stereotype-primed female arts students performing a mental rotation task. Despite having the same average math grades, a recent study found that female students report higher levels of trait math anxiety than male students, as assessed using experience sampling methods while students took a math test and attended math classes (Goetz, Bieg, Ludtke, Pekrun & Hall, 2013).. The same study did not reveal sex differences in state math anxiety. The authors concluded that discrepancies between trait and state anxiety partly accounted for students’ beliefs about their competence in math. Similar adverse effects of high trait anxiety (independent of sex) have been shown for visuospatial processing (Eysenck, Payne, & Derakshan, 2005). Although the underlying mechanisms are not fully understood, a recent systematic literature review (Eysenck, Deraksgan, Santos, & Calvo, 2007) suggested that anxiety increases attention to threat-related stimuli with negative consequences on processing efficiency. Specifically, the authors argued that anxiety negatively impacts on attentional control - the ability to resist interference from distracting external or internal stimuli (e.g., being afraid of confirming a negative stereotype). However, the authors also concluded that anxiety may not impair cognitive performance when it leads to the use of compensatory strategies, such as enhanced effort or increased use of processing resources (Eysenck et al., 2007, p. 336). This might explain the stereotype reactance effects found in the current study; the performance increase in a ‘female’ cognitive domain (i.e., word fluency) in stereotype-primed men. Together with the finding that self-ratings in men were generally higher (or at least the same for verbal fluency) than those in women, these findings suggest that self-confident participants, low in trait anxiety, may be generally less susceptible to stereotype threat. This observation has important implications for psychological interventions.

Practice of stress-reduction techniques and self-affirmation are only two promising avenues for intervention in women (Inzlicht & Ben-Zeev, 2003, Shapiro & Williams, 2012). Informing group members targeted by negative stereotypes about the effects of stereotype threat has been suggested as another strategy to buffer women's cognitive performance on stereotype-relevant tasks (Johns, Schmader & Martens, 2005). In this respect, the present paper might contribute to helping women break through the glass ceiling and enhance their performance in STEM disciplines. However, beyond that, it will likely require additional steps at the society and policy level to break the perceived correlation between gender and academic disciplines.

Limitation and future directions

The current study investigated the cognitive effects of stereotype priming in male and female students of arts (English literature and philosophy) and science disciplines (chemistry and engineering). Groups of science and arts students were not further divided by major. Although the current study did not reveal significant differences between majors within in each academic group (results not reported), future research could increase sample size and recruit participants from a wide range of disciplines. This would allow identification of (STEM) disciplines that are particularly prone to the effects of stereotype priming. Although the small sample size is a potential weakness, the sample sizes in the current study are similar to other studies in the stereotype threat literature. Also, the current study used only two cognitive tasks that have been shown to be particularly sex-sensitive. It would be interesting, however, to use a battery of (sex-sensitive) cognitive tasks that differ not only in cognitive domains (male or female), but also in task difficulty and in the cognitive sub-processes involved. For example, it is likely that stereotype threat occurred only in the mental rotation test, because this task is (a) particularly sex-sensitive, (b) more demanding than the word fluency task, and (c) requires more central executive processes, such as working memory and attentional control. To fully understand the mechanisms underlying stereotype-priming effects, future studies need to combine biological (i.e., endocrine and neural), psychological, and social factors within a single experimental approach.

The current study suggests that women who are less susceptible to gender stereotype threat, partly because of higher self-confidence in their cognitive abilities and/or lower trait anxiety, are more likely to perform better in a testing situation for which a negative gender stereotype exists. Future research is needed to investigate whether these findings also apply to real-world, job-relevant situations (e.g., job interviews, assessment center, etc.). It is important to note that the longitudinal link between the susceptibility to stereotype threat and later occupational status remains an open empirical question.

Acknowledgement

I thank Ailis Gavan, Sam Moore, Emily Quill, and Laura Savage for their assistance with data collection, and all participants for their cooperation.

References

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: a meta-analysis. *Sex Roles, 20*, 327-344.

Bargh, J. A. (2007). The automaticity of everyday life. In R. S. Wyer Jr. (Ed.), *Advances in* *social cognition* (Vol., pp. 1-61). Mahwah, NJ: Erlbaum.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230-244.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*, 174-181.

Casey, M. B., & Brabeck, M. M. (1989). Exceptions to the male advantage on a spatial task: Family handedness and college major as factors identifying women who excel. *Neuropsychologia, 27*, 689-696.

Crisp, R. J., Bache, L. M., Maitner, A. T. (2009). Dynamics of social comparison in counter-stereotype domains: Stereotype boost, not stereotype threat for women engineering majors. *Social influence, 4*, 171-184.

Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Archives of Sexual Behavior, 41*, 557-570.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attention control theory. *Emotion*, 7, 336-353.

Eysenck, M. W., Payne, S., & Derakshan, N. (2005). Trait anxiety, visuospatial processing, and working memory. *Cognition and Emotion*, 19, 1214-1228.

Goetz, T., Bieg, M., Ludtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science*, 24, 2079-2087.

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. Lawrence Erlbaum Associates, Publishers, Mahawah, NJ.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A.

(2007). The science of sex difference in science and mathematics. *Psychological Science in the Public Interest, 8*, 1-51.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, 12, 229-246.

Halpern, D. F., & Tan, U. (2001). Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences. *Brain and Cognition, 45*, 392-414.

Hausmann, M., Schoofs, D., Rosenthal, H. E. S., & Jordan, K. (2009). Interactive effects of sex hormones and gender stereotypes on cognitive sex differences – a psychobiosocial approach. *Psychoneuroendocrinology, 34*, 389-401.

Heil, M., Jansen, P., Quaiser-Pohl, C., & Neuburger, S. (2012). Gender-specific effects of artificially induced gender beliefs in mental rotation. *Learning and Individual Differences*, *22*, 350-353.

Hirnstein, M., Andrews, L. C., & Hausmann, M. (2014). Gender-stereotyping and cognitive sex differences in mixed- and same-sex groups. *Archives of Sexual Behavior*, 43, 1663-1673.

Hirnstein, M., Freund, N., & Hausmann, M. (2012). Gender stereotyping enhances verbal fluency performance in men (and women). *Zeitschrift für Psychologie, 220*, 70-77.

Horn, W. (1962). *Leistungsprüfsystem (LPS)*. Hogrefe, Göttingen.

Hyde, J.S., & Linn, M.C. (1988). Gender differences in verbal abilities: a meta-analysis. *Psychological Bulletin, 104*, 53-69.

Hyun, J.-S., & Luck, S. J. (2007). Visual working memory as the substrate for mental rotation. *Psychonomic Bulletin and Review, 14*, 154-158.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*, 175–179.

Jordan, K. (2010). The brain between sex and gender: Women and men from a neuroscientific perspective. In I. Klinge & C. Wiesemann (Ed.), *Gender and sex in biomedicine: theories, methodologies, results*, (79–99). Göttingen: Universitätsverlag Göttingen.

Josephs, R. A., Newman, M. L., Brown, R. P., & Beer, J. M. (2003). Status, testosterone, and human intellectual performance: Stereotype threat as status concern. *Psychological Science, 14*, 158-163.

Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95*, 796-805.

Kray, L. J., Thompson, L., & Galinsky, A. D. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology, 80*, 942-958.

Linn, M. C., & Petersen, A. C., (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479-1498.

Lubinski, D. (2010). Spatial ability and STEM: a sleeping giant for talent identification and development. *Personality and Individual Differences*, 49, 344–351.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing. *Behavior Genetics*, *23*, 337-341.

McGlone, J. (1980). Sex-differences in human-brain asymmetry: A critical survey. *Behavioral and Brain Sciences*, *3*, 215-227.

McKeever, W. F. (1995). Hormone and hemisphericity hypotheses regarding cognitive sex differences: Possible future explanatory power, but current empirical chaos. *Learning and Individual Differences, 7*, 323-340.

Moè, A. (2012). Gender difference does not mean genetic difference: Externalizing improves performance in mental rotation. *Learning and Individual Differences*, *22*, 20-24.

Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences, 19*, 21–27.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*, 369–377.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18*, 879-885.

Nazareth, A., Herrera, A., & Pruden, S. M. (2013). Explaining sex differences in mental rotation: role of spatial activity experience. *Cognitive Processing*, 14, 201-204.

North, S. (2005). Different values, different skills? A comparison of essay writing by students from arts and science backgrounds. *Studies in Higher Education, 30*, 517-533.

Nguyen, H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*, 1314-1334.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29*, 782-789.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A Redrawn Vandenberg and Kuse Mental Rotation Test: Different versions and factors that affect performance. *Brain and Cognition, 28*, 39-58.

Picho, K., Rodriquez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: a meta-analysis. *Journal of Social Psychology, 153*, 299-333.

Richardson, J. T. E. (1994). Gender differences in mental rotation. *Perceptual and Motor Skills, 78*, 435-448

Rydell, R. J., Van Loo, K. J., & Boucher, K. L. (2014). Stereotype threat and executive functions: Which functions mediate different threat-related outcomes. *Personality and Social Psychology Bulletin*, in press.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*, 194-201.

Schmader, T., Johns, M., & Forbes, C. (2005).An integrated model of stereotype threat effects on performance. *Psychological Review, 115*, 336-356.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 440-452.

Shapiro, J., & Williams, A. M. (2012). The role of stereotype threat in undermining girls' and women's performance and interest in STEM fields. *Sex Roles, 66*, 175-183.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.

Sherman, J. (1967). Problem of sex differences in space perception and aspects of intellectual functioning. *Psychological Review, 74*, 290-299.

Shih, M., Ambady, N., Richeson, J.A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology, 38*, 638-647.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4-28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613-629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.

Summers, L. H. (2005, January 14). Remarks at NBER Conference on Diversifying the Science & Engineering Workforce, Cambridge, M.A. Retrieved June 8, 2013, from http://www.harvard.edu/president/speeches/summers_2005/nber.php

Swim, J. K. (1994). Perceived versus meta-analytic effect size: an assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology, 66*, 21-36.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of 3-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*, 599-604.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250-270.

Wai, J., Lubinski, D., & Benbow, C. P. (2009) Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101, 817–835.

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456-467.

Werhun, C. D. (2007). The limitations of stereotype threat: not all math and science women are threatened by stereotypes. (Doctoral dissertation). University of Toronto, Toronto, Canada.

Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin and Review, 13*, 813-819.

Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognitive and Affective Neuroscience, 2*, 12-19.

Zimmer, H. D. (2008). Visual and spatial working memory: from boxes to networks. *Neuroscience and Biobehavioral Reviews*, 32, 1373-1395.